Combined Data-Driven Biomedical Outcome Prediction and Interaction Network Inference from Molecular Profiling Data

The International Science Forum on Computational Toxicology

May 21–23, 2007, U.S. EPA Main Facility, Research Triangle Park, NC

Roland Somogyi, Ph.D.

May 22, 2007

rsomogyi@biosystemix.com

www.biosystemix.com

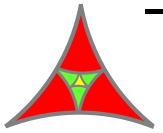
(613)-376-3126

Data-driven “reverse engineering” analysis directly links measurements to insight

- Computational solutions can be achieved that are closely and directly tied to observed facts
- Molecular activity patterns are the result of the molecular interaction networks
- Ergo: Molecular interaction networks may be “reverse engineered" from activity patterns
- Furthermore, reverse engineerig approaches, as applied to molecular activity patterns of toxicological responses, may provide key, **objective** insights into processes of toxicity

Three fundamental approaches

- PIA – Predictive Interaction Networks
  - Does gene A interact with gene B in determining a response or cell/tissue type?
  - Directly identifies the gene interactions that determine biological outcomes
- CFA – CoFluctuation Analysis
  - Is the profile of gene A correlated to gene B?
  - Tells us about shared regulation, but not how genes interact to determine biological outcomes
- TNA – Temporal Network Analysis
  - Does the activity of gene A predict gene B?
  - Tells us about how genes regulate each other, but not how interactions determine outcomes



PIA – Predictive Interaction Analysis

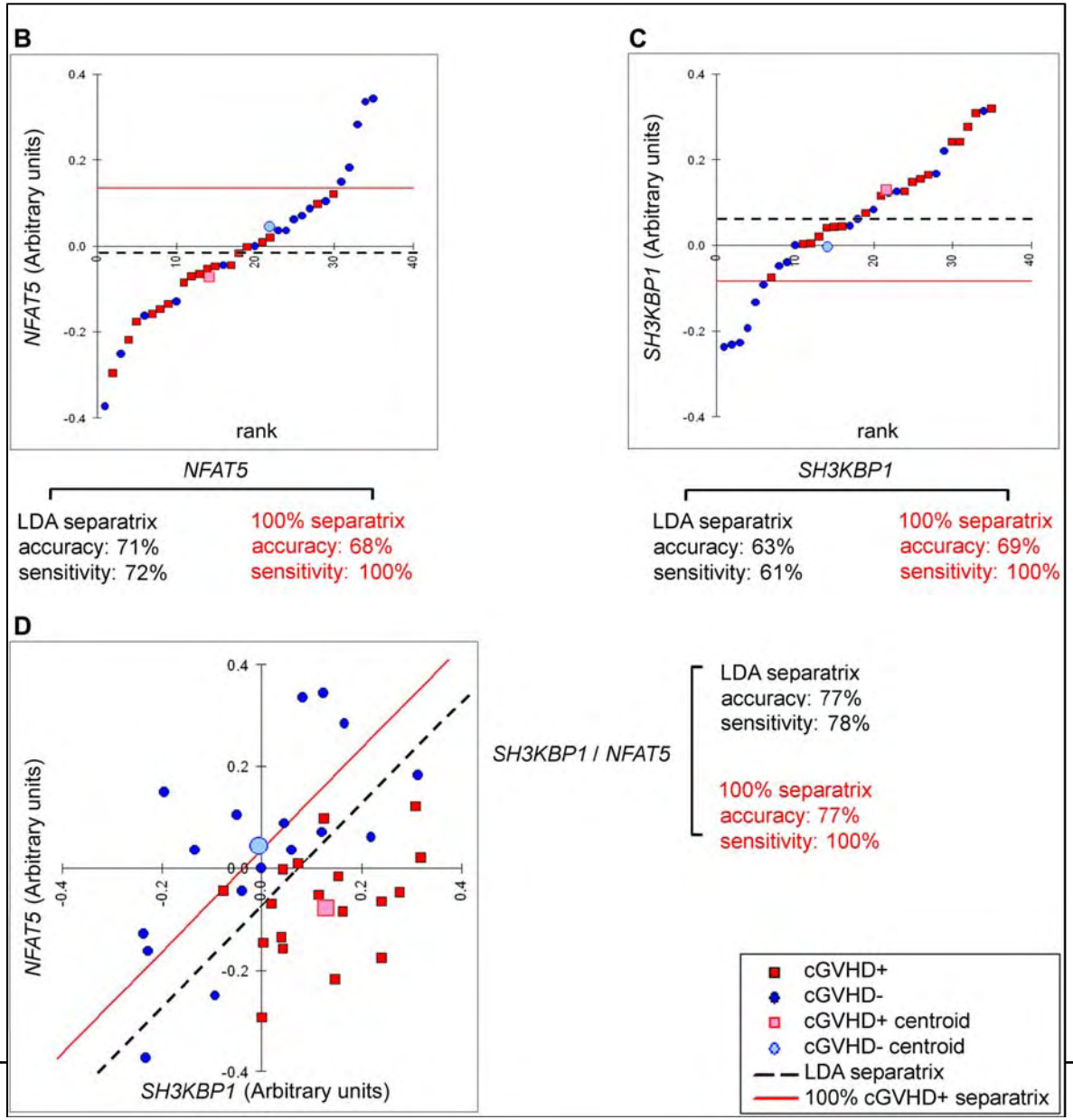
Competitive Predictive Interaction Analysis

- Gene pair expression measurements are reduced to a single variable, **v**
  - **x/y ~ outcome**, where x represents gene X (R/G) and y represents gene Y (R/G)
  - **logx - logy ~ outcome**
  - **we define: v = logx – logy**
    - Note: In the analysis below, x then will refer to logx, and y then will refer to logy
- ANOVA provides p-values to assess separation of outcome class-specific distributions
- The true value of a GP (gene **pair**) model can only be demonstrated if it outperforms predictive models based on its constituent genes

Synergistic Predictive Interaction Analysis

- Gene pair expression measurements are reduced to a single variable, **u**
  - **x*y ~ outcome**, where x represents gene X (R/G) and y represents gene Y (R/G)
  - **logx + logy ~ outcome**
  - **we define: u = logx + logy**

PIA gene pair model best predicts GVHD (graft vs. host disease), compared to single genes



- PIA gene pair p-value is 1000x better than the best single gene

Baron et al. (2007) Prediction of graft-versus host disease in humans by donor gene-expression profiling. PLoS Med 4(1): e23.

Similarity Analysis for identification of key features in vaccination time series

CONCORDANCE OF MEANS (class-class comparison)

	t000	t003	t007	t010	t014	t028	t060	t180
t000	1.00	0.84	0.81	0.97	0.96	0.97	0.96	0.96
t003	0.84	1.00	0.97	0.83	0.81	0.82	0.83	0.81
t007	0.81	0.97	1.00	0.81	0.78	0.80	0.80	0.79
t010	0.97	0.83	0.81	1.00	0.96	0.97	0.95	0.95
t014	0.96	0.81	0.78	0.96	1.00	0.97	0.94	0.97
t028	0.97	0.82	0.80	0.97	0.97	1.00	0.96	0.96
t060	0.96	0.83	0.80	0.95	0.94	0.96	1.00	0.95
t180	0.96	0.81	0.79	0.95	0.97	0.96	0.95	1.00

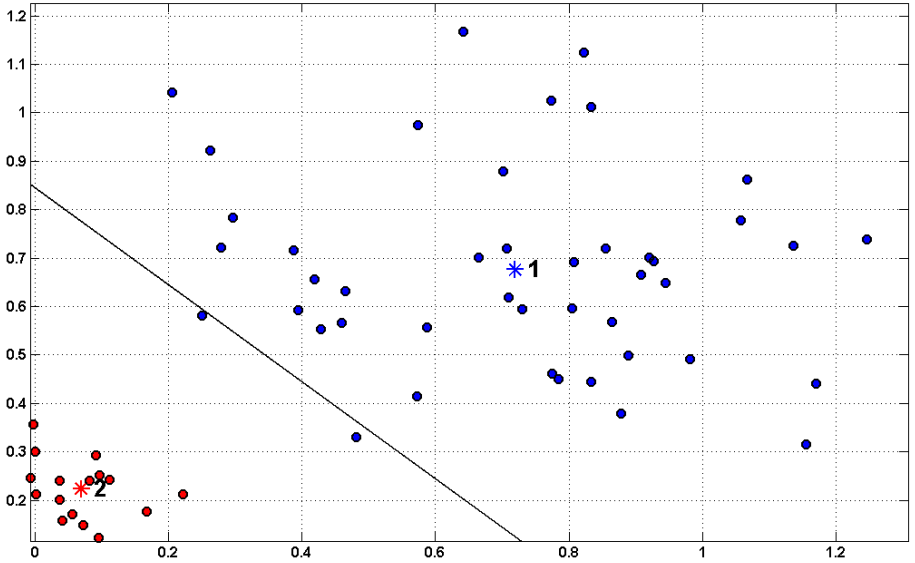
MEAN OF CONCORDANCE (chip-chip comparison)

	t000	t003	t007	t010	t014	t028	t060	t180
t000	0.64	0.65	0.62	0.73	0.72	0.74	0.74	0.72
t003	0.65	0.67	0.75	0.64	0.62	0.64	0.64	0.62
t007	0.62	0.75	0.63	0.61	0.59	0.61	0.62	0.59
t010	0.73	0.64	0.61	0.63	0.72	0.74	0.72	0.72
t014	0.72	0.62	0.59	0.72	0.62	0.73	0.71	0.72
t028	0.74	0.64	0.61	0.74	0.73	0.62	0.74	0.73
t060	0.74	0.64	0.62	0.72	0.71	0.74	0.60	0.72
t180	0.72	0.62	0.59	0.72	0.72	0.73	0.72	0.60

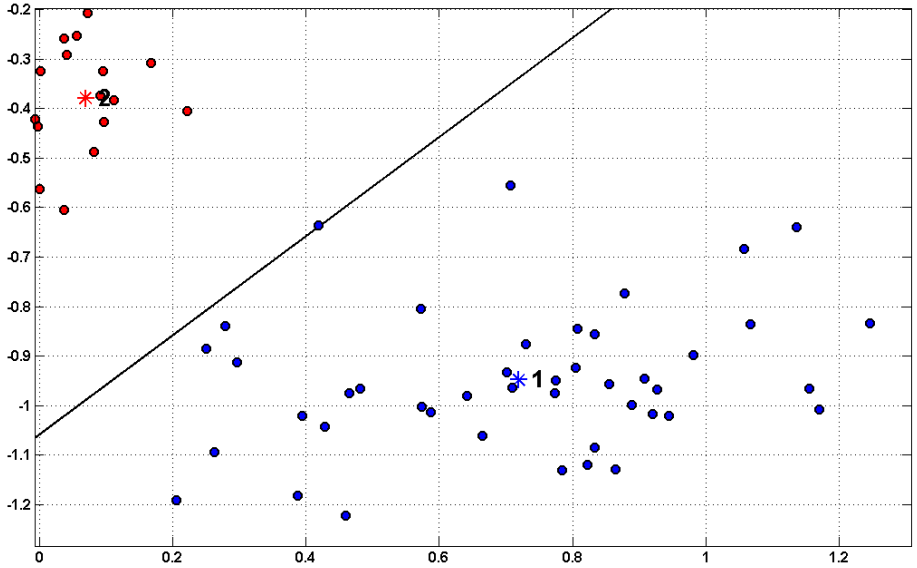
- The major time point differences are between t003 and t007 vs. the remaining time points

PIA of vaccination response shows strong gene interactions in outcome discrimination

Best synergistic model
log (gene x * gene y)

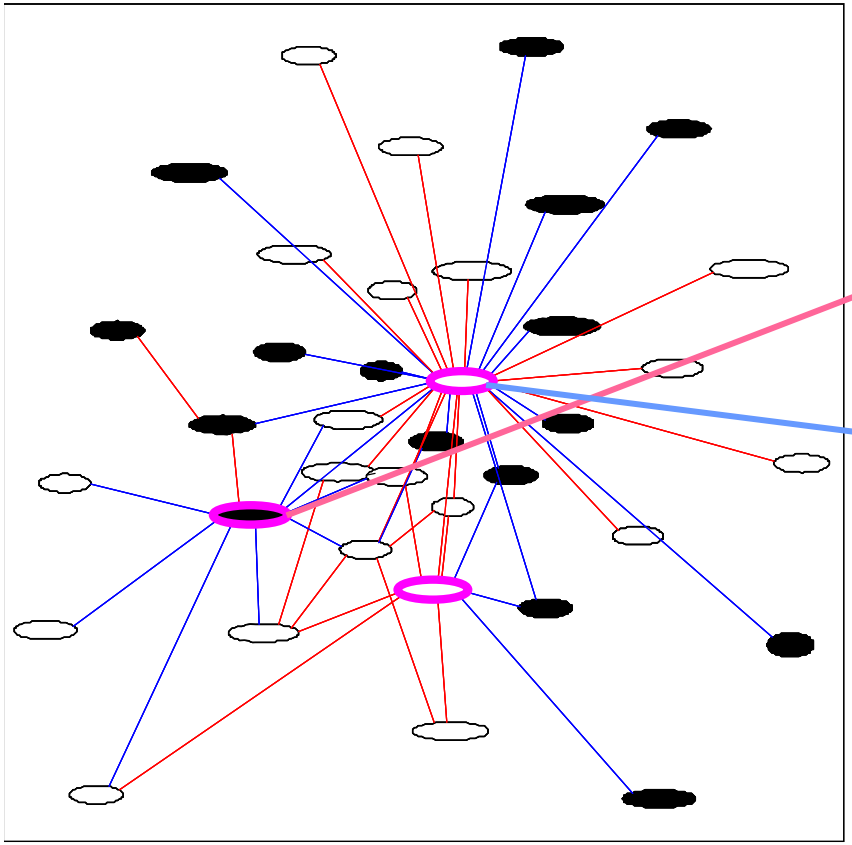


Best competitive model
log (gene y / gene x)

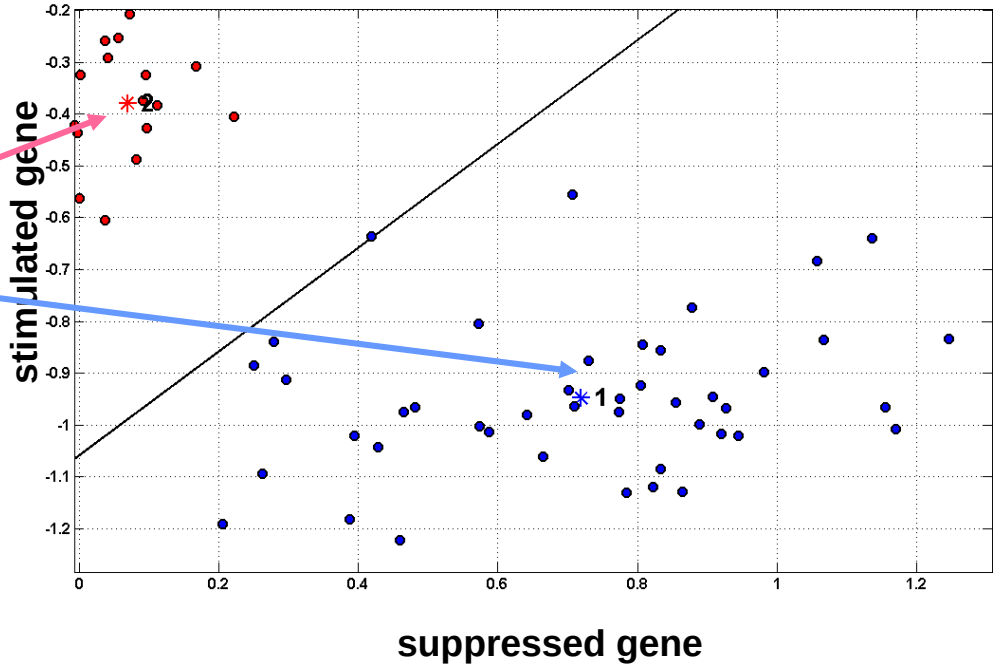


Best model	Gene X	Gene Y	GP SPIA log(10) p-value	GP CPIA log(10) p-value	SG X log(10) p-value	SG Y log(10) p-value	GP SPIA accuracy	GP CPIA accuracy	SG X accuracy	SG Y accuracy	GP log(10) p-value gain	GP accuracy gain
SPIA (synergistic)	Protein kinase in cell cycle control	Tumor-related gene	**-27.3**	-2.7	-19.8	-18.0	**97%**	67%	90%	90%	**7.6**	**7%**
CPIA (competitive)	Protein kinase in cell cycle control	Neurotransmitter receptor	-0.7	**-30.6**	-19.8	-17.1	60%	**98%**	90%	95%	**10.8**	**3%**

PIA gene pairs form a network of interactions that determine vaccination response outcome



Competitive model



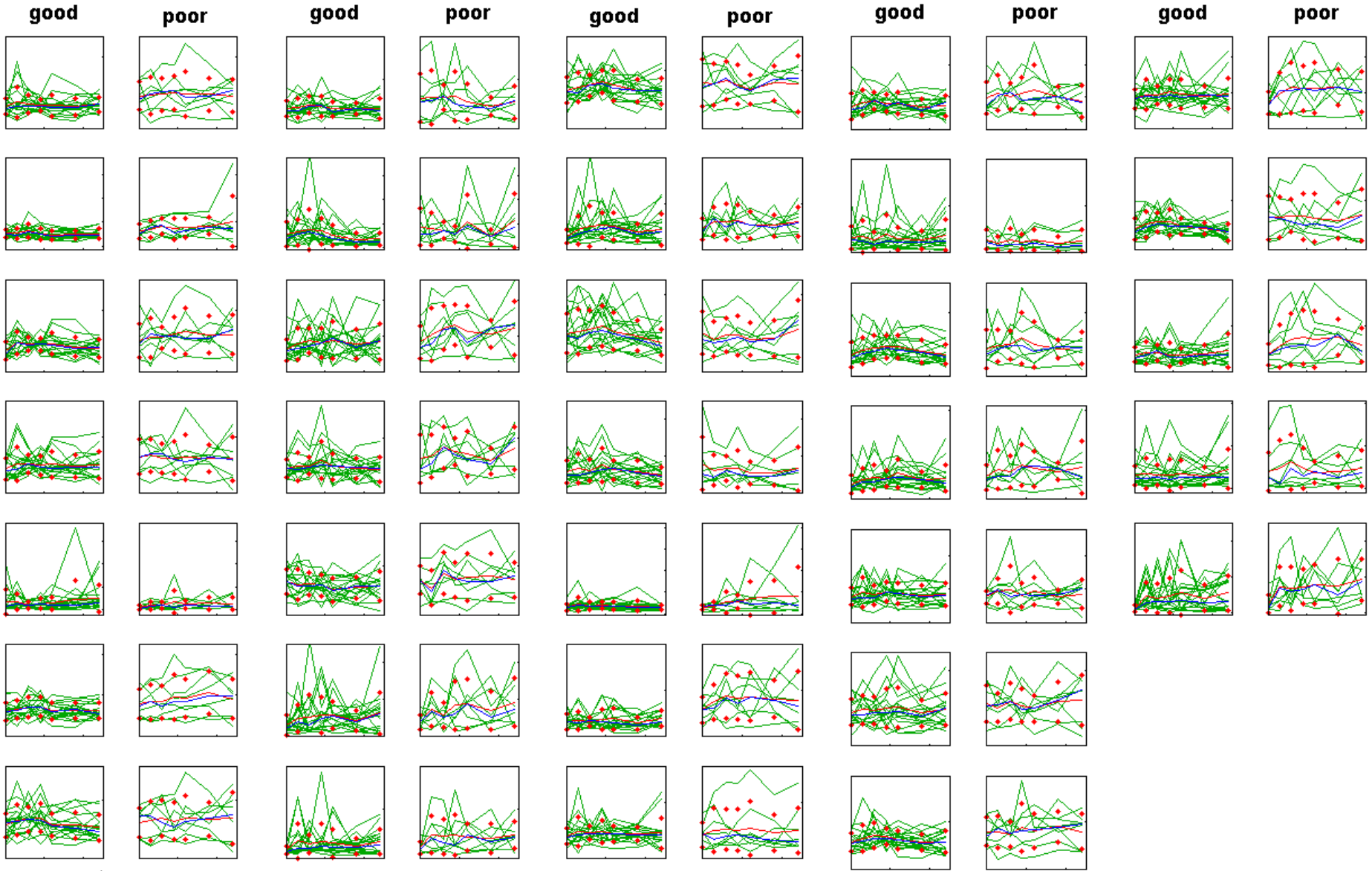
Red arrows:
Syngergistic PIA interaction

Blue connections:
Competitive PIA interactions

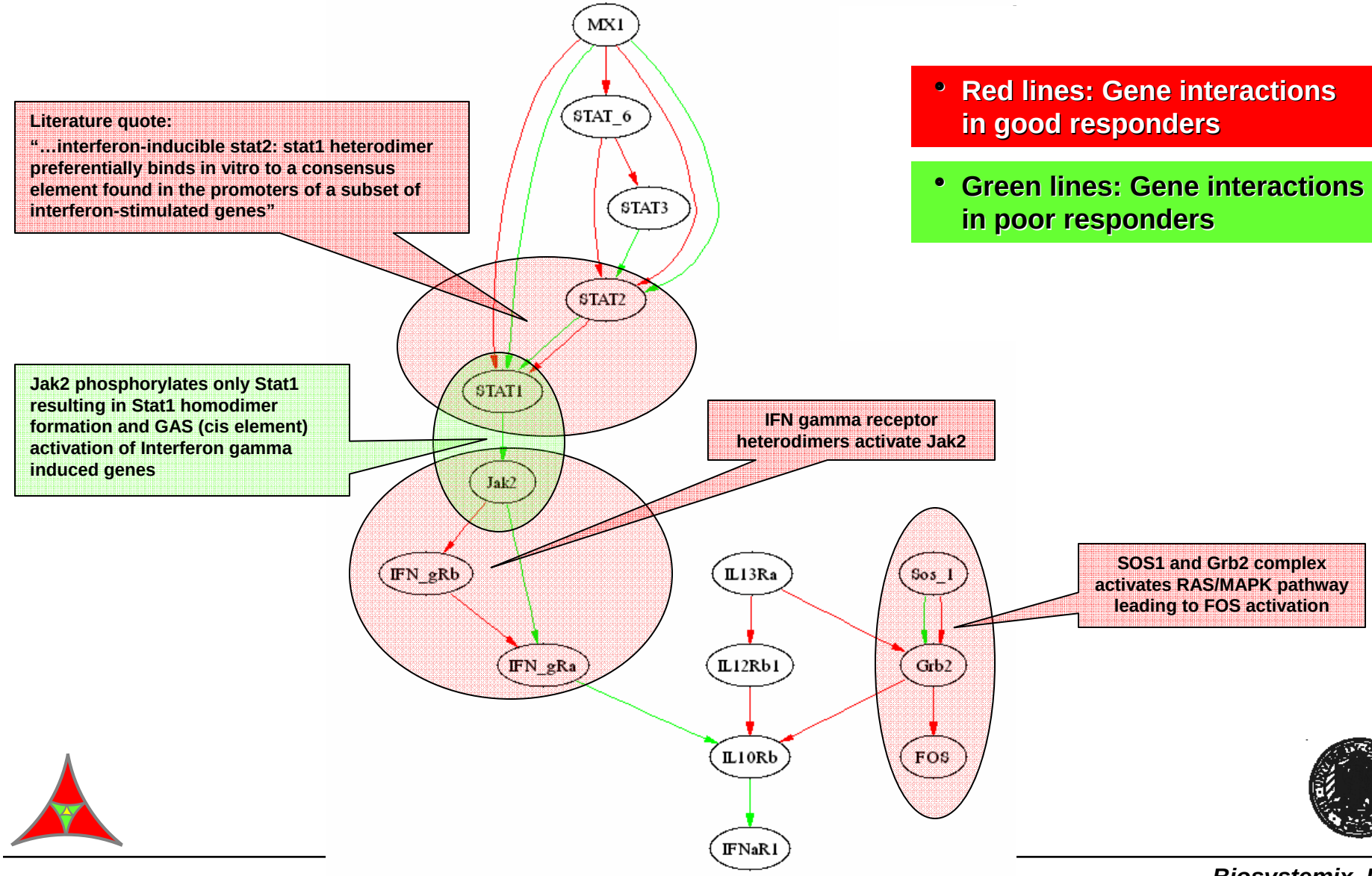
Red: D3D7 time points
Blue: B (baseline) time points

CFA – CoFluctuation Analysis

Expression time series over seven time points for multiple sclerosis patients are highly variable



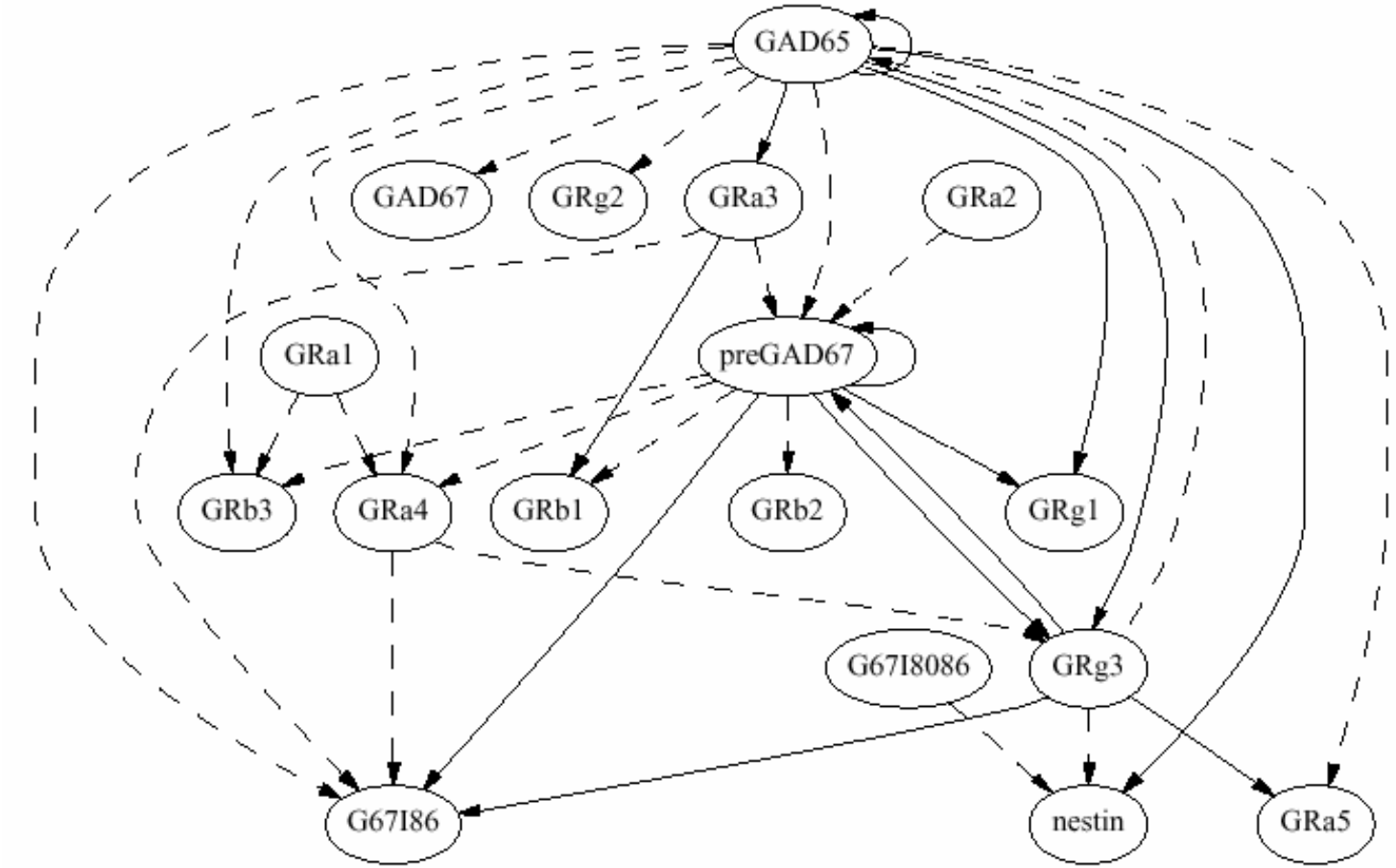
CFA reconstructs signaling pathways directly from clinical multiple sclerosis gene expression data



TNA – Temporal Network Analysis

Reverse Engineering the GABAergic Gene Network from Expression Time Series Data

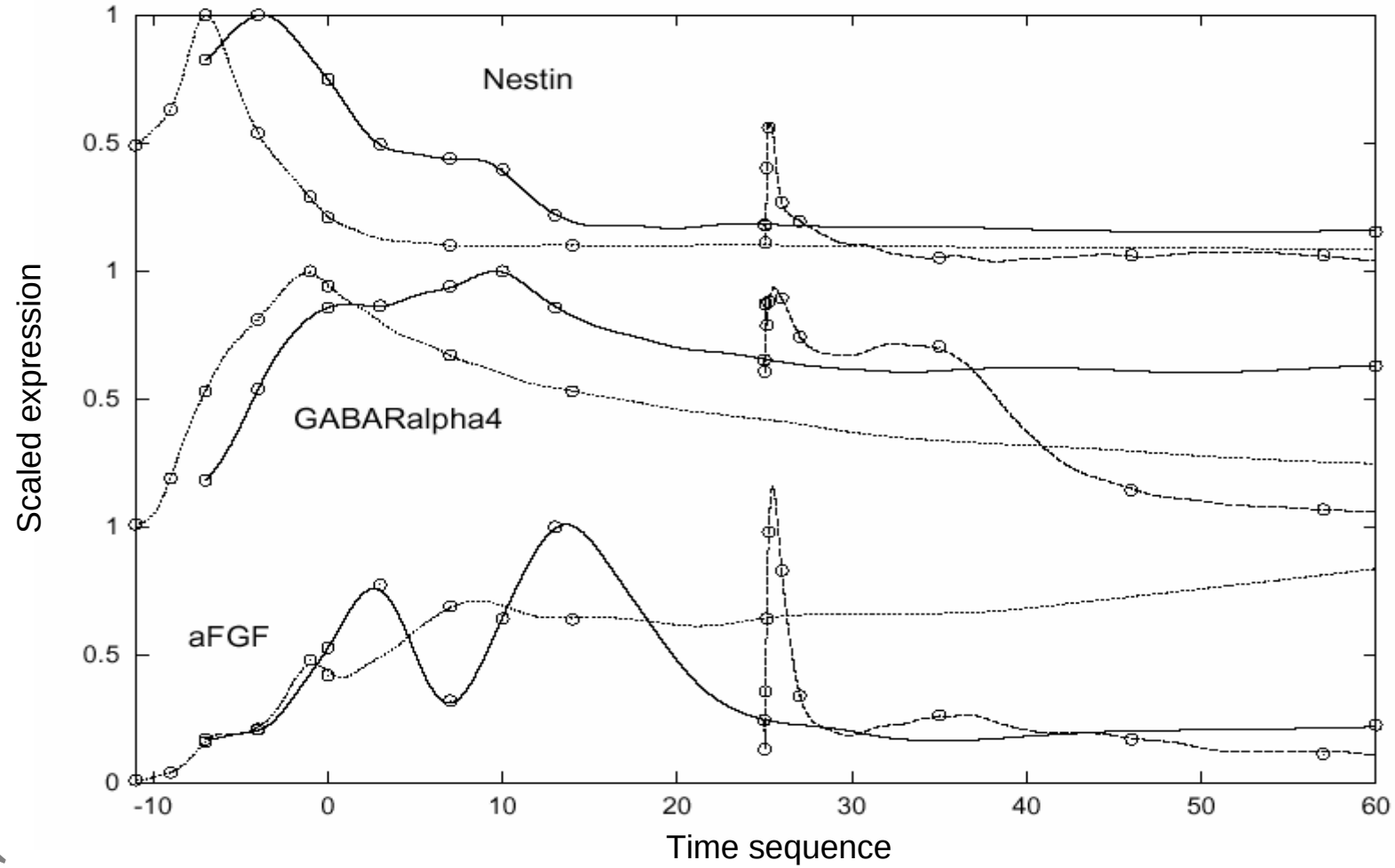
Connectivity inferred computationally from the data using linear differential equation models



Ref: P. D'Haeseleer, X. Wen, S. Furhman, R. Somogyi, *Proc. Pacific Symp. Biocomp.* 4, 41-52 (1999)

In silico Prediction of Experimental Gene Expression using a Reverse-Engineered Gene Network Model

In addition to providing a wiring diagram, TNA accurately simulates original measurement data



Ref: P. D'Haeseleer, X. Wen, S. Furhman, R. Somogyi, *Proc. Pacific Symp. Biocomp.* 4, 41-52 (1999)

Biosystemix Discovery Partnership Areas

- Clinical studies
  - Personalized molecular medicine (development of novel molecular diagnostics)
  - Functional discovery (molecular interaction network inference)
- Laboratory models
  - **Toxicity** and efficacy studies
  - Complex dosing and time series studies
  - Complexity of perturbation data provides support for gene network reverse engineering

Biosystemix Discovery Partnerships

- University of Montreal &
  Institute for Research in Cancer and Immunology
  - Prediction of GVHD (graft vs. host disease)
  - Understanding of HIV resistance
- Queens University &
  Ontario Institute for Cancer Research
  - Predicting clinical outcomes in Follicular Lymphoma
- University of Manitoba
  - Understanding of HIV resistance
- UCSF, Department of Neurology
  - Predicting drug response in multiple sclerosis
- University of Michigan
  - Inference of pathways involved in toxicity